



Data Warehousing

Politecnico di Torino

Food delivery

Problem specifications

A team of data scientists has been charged with a set of analyses on the data collected by a food delivery company. For the analysis they have to consider the restaurants associated to deliveries. For each restaurant the name and the city where it is located are known. Moreover, restaurants can be characterized by one or more categories (e.g. “Indian”, “Italian”, “Pizza”, ..., “Hamburger”, “Sandwiches”). The number of categories is 10 and their complete list is known. Based on the restaurant, the order associated to the delivery is characterized by a type which can assume the values “Transport or sending”, “Partner order”, “Non-partner Order”. For example the partner orders are associated to restaurants which are partner of the food delivery company.

It is necessary to consider the neighborhood of destination for the delivery and the payment method (e.g. “Bancomat”, “Credit card”, ...). The analysis must be carried out considering the date and the time slot of the deliveries. Each time slot lasts exactly 1 hour (e.g. “9-10”, “10-11”).

Each delivery is associated to a rider. For the analyses it is necessary to know the mean of transport used by the rider (e.g. “bike”, “car”).

The company is interested in the statistics on the average number of kilometers run by riders, the average delivery time (in minutes) and the average revenue for delivery. The analysis must be carried on based on:

- restaurant name, restaurant category, city, province, region
- order type
- delivery neighborhood, city, province, region
- payment method
- mean of transport of the rider
- delivery date, month, 2-months, trimester, 4-months, 6-months, year, working day
- time slot

Design

Design the data warehouse to address the specifications and to efficiently answer to all the provided frequent queries. Draw the conceptual schema of the data warehouse and the logical schema (fact and dimension tables).

Query

Write the following frequent queries using the extended SQL language

- a) Consider the orders with type “Partner order”. Separately for mean of transport and trimester, analyze: the average delivery time, the average number of deliveries per time slot (the average number of deliveries made in an hour), assign a rank to the trimesters based on decreasing number of kilometers run on average in a minute, separately by mean of transport.
- b) Consider the restaurants which have “pizza” among the associated categories. Carry out the analysis separately for payment method, delivery city and month. Analyze: the cumulative monthly revenue from the beginning of each trimester, the average revenue per delivery, the percentage of revenue with respect to the total revenue considering all the payment methods